

Moment estimators of the GEV distribution for the minima

Jose Raynal

Received: 21 October 2011 / Accepted: 12 July 2012 / Published online: 9 August 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract The moment (MOM) estimators for the parameters, quantiles and confidence limits, using the general extreme value distribution for the minima, is presented towards its application in low flow frequency analysis. The procedures to compute the parameters, design events (quantiles) for several return periods and their confidence limits are shown in the paper. Two measures of goodness of fit tests are contained in the paper to compare the proposed methodology with other models in competition. A full example of application is presented in the paper to show how easy is to apply the proposed methodology.

Keywords Low flow · Frequency analysis · Parameter estimation · Confidence limits · Method of moments

Introduction

A subject of paramount interest in planning and design of water works is that related with low flow frequency analysis. Due to the characteristic that design values have, given that they are linked to a return period or to an exceedance probability, the use of mathematical models known as probability-distribution functions is a must. Among the most widely used probability distribution functions for hydrological analysis, related with low flow frequency analysis, are (Kite 1988; Salas and Smith 1980; Rao and Hamed 2000; Raynal-Villasenor 2010):

1. Three parameters log-normal (LN3)
2. Pearson type III (PIII)
3. Extreme value type III (EVIIM)
4. General extreme value for the minima (GEVM)

The first three probability-distribution functions have been applied to low flow frequency analysis by Matalas (1963). Gumbel (1958) developed the theoretical grounds and hydrological applications for the extreme value type III distribution for the minima (EVIIM), the well-known Weibull distribution. This distribution has been applied since the first third of the twentieth century to the analysis of dynamic breaking strength of materials (Weibull 1939a, b). Kite (1988) provided with a computer program to estimate the parameters of the EVIIM distribution using the methods of moments (MOM) and maximum likelihood (ML). More recently, Lee and Kim (2008) used the two-parameter Weibull distribution with Bayesian Markov chain Monte Carlo and maximum likelihood estimates to assess the uncertainty of low frequency analysis. The estimation of ML parameters of EVIIM distribution has some difficulties when using the Newton–Raphson method as have been pointed out by Offinger (1996). Durrans and Tomic (2001) compared five methods of estimation of parameters for the log-normal distribution in fitting the lower tail of such distribution. Smakhtin (2001) made a review of 20 years of research results with regard to low flow hydrology. Yue and Wang (2004) studied the scaling of Canadian rivers to regionalize the low flows. Taha et al. (2008) presented a brief review of statistical models that are commonly used in the estimation of low flows both at sites with a reliable stream flow record and sites remote from data sources. Hao and Singh (2009) applied the maximum entropy method to the Burr III distribution and compared the results with the MOM, ML and probability-weighted moments (PWM); they found no differences on the

J. Raynal (✉)
Universidad de las Americas Puebla,
72810 San Andres Cholula, Puebla, Mexico
e-mail: joseraynal@hotmail.com

quantiles for small return period, the differences increased for large period returns. Iacobellis (2008) studied the evaluation of a flow duration curve with assigned a T -year return period with beta and complementary beta distributions.

The use of the general extreme value distribution for the minima (GEVM) with moment estimators for the parameters, quantiles and confidence limits are proposed in the paper. A complete example of application of the proposed methodology is contained in the paper, through the application of common spreadsheets framework provided by Excel® (Excel is a registered trademark of Microsoft Corporation, Inc.). The results are compared with the other three distribution function mentioned before.

Probability distribution and density functions of the GEVM

The probability-distribution function of the GEVM distribution for the minima is, Raynal-Villasenor and Douriet-Cardenas (1994):

$$\Pi(x) = \exp\left\{-[1 - \beta(\omega - x)/\alpha]^{1/\beta}\right\} \quad (1)$$

where ω , α and β are the location, scale and shape parameters, respectively. $\Pi(x)$ is the probability-distribution function of the random variable x and for the case of low flow frequency analysis is equal to the exceedance probability, $\Pr(X > x)$. The scale parameter must meet the condition that $\alpha > 0$. The domain of variable x in GEVM distribution is as follows:

$$1. \text{ For } \beta < 0: \\ -\infty < x \leq \omega - \alpha/\beta \quad (2)$$

$$2. \text{ For } \beta > 0: \\ \omega - \alpha/\beta \leq x < \infty \quad (3)$$

The probability density function for the GEVM distribution is (Raynal-Villasenor and Douriet-Cardenas 1994):

$$\pi(x) = \frac{1}{\alpha} \exp\left\{-[1 - \beta(\omega - x)/\alpha]^{1/\beta}\right\} [1 - \beta(\omega - x)/\alpha]^{1/\beta - 1} \quad (4)$$

where $\pi(x)$ is the probability-density distribution of random variable x .

Moment estimators for the parameters of the GEVM distribution

The moment estimators for the parameters of the GEVM distribution have the following expressions:

1. Location parameter:

$$\hat{\omega} = \hat{A} + \frac{\hat{\alpha}}{\hat{\beta}} \quad (5)$$

$$\hat{A} = \hat{\mu} - \frac{\hat{\alpha}}{\hat{\beta}} \Gamma(1 + \hat{\beta}) \quad (6)$$

2. Scale parameter:

$$\hat{A} = \hat{\mu} - \frac{\hat{\alpha}}{\hat{\beta}} \Gamma(1 + \hat{\beta}) \quad (7)$$

$$\hat{B} = \left(\frac{\hat{\sigma}^2}{\sigma_z^2}\right)^{1/2} = \frac{\hat{\sigma}}{\sigma_z} \quad (8)$$

$$\sigma_z^2 = \Gamma(1 + 2\beta) - \Gamma^2(1 + \beta) \quad (9)$$

3. Shape parameter:

For $\beta < 0$ and $-19.0 < \hat{\gamma} \leq -1.1396$:

$$\hat{\beta} = 0.24662 + 0.286678\hat{\gamma} + 0.072454\hat{\gamma}^2 \\ + 0.010176\hat{\gamma}^3 + 0.000816\hat{\gamma}^4 + 0.000037\hat{\gamma}^5 \quad (10)$$

For $\beta > 0$ and $-1.1396 \leq \hat{\gamma} < 11.35$:

$$\hat{\beta} = 0.279434 - 0.333535\hat{\gamma} + 0.048305\hat{\gamma}^2 \\ + 0.024414\hat{\gamma}^3 + 0.003765\hat{\gamma}^4 - 0.000263\hat{\gamma}^5 \quad (11)$$

where x_0 , α and β are the location, scale and shape parameters of the GEVM distribution. $\Gamma(\cdot)$ is the complete Gamma function.

Design values for the GEVM distribution

The design values (quantiles) for the GEVM distribution can be obtained by inverting the GEVM distribution function:

$$Q_T = \omega + \frac{\alpha}{\beta} \left\{ \left[-\text{Ln} \left(1 - \frac{1}{T_r} \right) \right]^\beta - 1 \right\} \quad (12)$$

where Q_T are the design values and T_r is the return period associated with such design values.

Confidence limits for the design values of the GEVM distribution

The moment confidence limits for the GEVM distribution are been computed through the following formula:

$$x_1 = Q_T \pm z_\alpha S_T \quad (13)$$

where x_1 is the confidence limit (lower or upper confidence limit), Q_T is the design value, z_α is a standard normal value

corresponding to a confidence level of α , and S_T is the standard deviation of the estimates. The form of such standard deviation is:

$$S_T^2 = \frac{\mu_2}{N} \left\{ 1 + K_T \hat{\gamma} + \frac{K_T^2}{4} (\hat{\kappa} - 1) + \frac{\partial K_T}{\partial \hat{\gamma}} \left[2\hat{\kappa} - 3\hat{\gamma}^2 - 6 + K_T \left(\hat{\lambda}_1 - \frac{6\hat{\gamma}\hat{\kappa}}{4} - \frac{10\hat{\gamma}}{4} \right) \right] \right\} \quad (14)$$

where S_T^2 is the variance of the estimates, μ_2 is the sample variance, N is the sample size, K_T is the frequency factor, γ is the skewness coefficient, κ is the kurtosis coefficient and λ_1 is a function of moments. Then, the frequency factor is:

$$K_T = B_K \left[\left(-\text{Ln} \left(1 - \frac{1}{T_r} \right) \right)^\beta - A_K \right] \quad (15)$$

$$A_K = \Gamma(1 + \beta) \quad (16)$$

$$B_K = \frac{1}{[\Gamma(1 + 2\beta) - \Gamma^2(1 + \beta)]^{1/2}} \quad (17)$$

$$\frac{\partial K_T}{\partial \gamma} = \left(\frac{\partial K_T}{\partial \beta} \right) \left(\frac{\partial \beta}{\partial \gamma} \right) \quad (18)$$

$$\frac{\partial K_T}{\partial \beta} = \left\{ \frac{[y^\beta \text{Ln}(y) - G_1 P_1] - \frac{1}{2} [y^\beta - G_1] [G_2 - G_1^2]^{-1} [G_2 P_2 - 2G_1^2 P_1]}{[G_2 - G_1^2]^{1/2}} \right\} \quad (19)$$

$$\frac{\partial \gamma}{\partial \beta} = \left\{ \frac{[G_3 P_3 + 3G_1 (P_1 (G_1^2 - G_2) - G_2 P_2)] - \frac{3}{2} [G_2 - G_1^2]^{-1} [G_3 - 3G_1 G_2 + 2G_1^3] [G_2 P_2 - 2G_1^2 P_1]}{[G_2 - G_1^2]^{3/2}} \right\} \quad (20)$$

$$y = -\text{Ln} \left(\frac{1}{T_r} \right) \quad (21)$$

$$G_r = \Gamma(1 + r\beta) \quad (22)$$

$$P_r = \psi(1 + r\beta) \quad (23)$$

where $\psi(\cdot)$ is the di-gamma function.

Goodness of fit tests for the parameters of the GEVM distribution

The two goodness of fit tests considered in this paper are:

1. Standard error of fit, SEF, Kite (1988)

$$\text{SEF} = \left[\frac{\sum_{i=1}^N (x_i - y_i)^2}{(N - n_p)} \right]^{1/2} \quad (25)$$

where x_i are the descending-ordered historical values of the sample, y_i are the values produced by the distribution function corresponding to the same return periods of the historical values, N is the sample size, and n_p is the number of parameters of the distribution function, in this case, $n_p = 3$.

2. Mean absolute relative deviation, MARD, Jain and Singh (1987)

$$\text{MARD} = \frac{100}{N} \sum_{i=1}^N \left| \frac{(x_i - y_i)}{x_i} \right| \quad (26)$$

Numerical example

The gauging station Villalba is located in the San Pedro River in Northwestern Mexico and has been selected to analyze its sample of annual one-day low flows, using the

$$\frac{\partial K_T}{\partial \gamma} = \frac{[G_2 - G_1^2] \{ [y^\beta \text{Ln}(y) - G_1 P_1] \} - \frac{1}{2} [y^\beta - G_1] [G_2 P_2 - 2G_1^2 P_1]}{\left\{ [G_3 P_3 + 3G_1 (P_1 (G_1^2 - G_2) - G_2 P_2)] - \frac{3}{2} [G_2 - G_1^2]^{-1} [G_3 - 3G_1 G_2 + 2G_1^3] [G_2 P_2 - 2G_1^2 P_1] \right\}} \quad (24)$$



Fig. 1 Location of gauging station Villalba, Mexico

	A	B
1	LOW-FLOW FREQUENCY ANALYSIS	
2	Gauging Station: Villalba, Mexico (1939-1991)	
3		
4	<i>Descriptive Statistics</i>	
5	Mean	0.3306
6	Standard Error	0.0201
7	Median	0.3280
8	Mode	0.4850
9	Standard Deviation	0.1465
10	Sample Variance	0.0214
11	Kurtosis	0.1906
12	Skewness	0.5853
13	Range	0.6250
14	Minimum	0.0790
15	Maximum	0.7040
16	Sum	17.5214
17	Count	53.0000
18	Largest(1)	0.7040
19	Smallest(1)	0.0790
20	Confidence Level(95.0%)	0.0404

Fig. 2 Data statistics for gauging station Villalba, Mexico

GEVM distribution with the MOM method of estimation of its parameters, design values and confidence limits.

The geographical location of gauging station Villalba, Mexico is shown in Fig. 1.

The first step in the computations is to obtain basic statistics of the one-day low flow sample and such statistics have been obtained by the application of common spreadsheets framework provided by Excel® (Excel is a registered trademark of Microsoft Corporation, Inc.), they are shown in Fig. 2.

The parameters, the goodness of fit measures, and design values and its confidence limits obtained through the use of the application of common spreadsheets framework provided by Excel® (Excel is a registered trademark of Microsoft Corporation, Inc.), they are shown in Fig. 3.

The comparison between the histogram of flood data and the theoretical probability-density function is shown in Fig. 4. Figure 5 shows the empirical and theoretical frequency curves for the MOM estimation of parameters for the GEVM, PIII, EVIIM and LN3 distributions to the 1-day low flow sample of gauging station Villalba, Mexico. In Fig. 6, it is shown a graphical representation of the MOM method of estimation for the design values and their confidence limits. All the figures mentioned before have been obtained through the use of the application of common spreadsheets framework

	A	B	C	D	E	F	G	H	I
1	LOW-FLOW FREQUENCY ANALYSIS								
2	Gauging Station: Villalba, Mexico (1939-1991)								
3	General Extreme Value (Minima)								
4	Method: Moments								
5									
6	Estimation of Parameters and Goodness of Fit Tests								
7	$1+\beta =$	1.4965	$1+2\beta =$	1.9930	$1+3\beta =$	2.4895			
8	$\Gamma(1+\beta) =$	0.8861	$\Gamma(1+2\beta) =$	0.9971	$\Gamma(1+3\beta) =$	1.3196			
9	$\psi(1+\beta) =$	0.0332	$\psi(1+2\beta) =$	0.4183	$\psi(1+3\beta) =$	0.6980			
10	$\mu_z =$	0.8861	$\sigma_z =$	0.4603	$\lambda_1 =$	0.0354			
11	A =	0.0486	B =	0.3182	$\lambda_2 =$	0.0425			
12	Location Parameter =		0.3668	m^3/s					
13	Scale Parameter =		0.1580	m^3/s					
14	Shape Parameter =		0.4965						
15	SEF =		0.02	MARD =		4.36			
16									
17	Design Values and Confidence Limits								
18	Return Period	Exceedance		Frequency		Squared Root	Two-sided	Design	Two-sided
19		Probability		Factor		Stand. Error	95%	Values	95%
20	T_r	$\Pi(x)$	y	K_T	$\partial K_T / \partial \gamma$	S_T	Lower Limit	Q_T	Upper Limit
21	(years)					(m^3/s)	(m^3/s)	(m^3/s)	(m^3/s)
22	2	0.5	0.6931	-0.1141	-0.0643	0.0239	0.2671	0.3139	0.3607
23	5	0.2	1.6094	0.8265	0.0731	0.0177	0.1650	0.1997	0.2345
24	10	0.1	2.3026	1.3620	0.1814	0.0052	0.1426	0.1527	0.1629
25	20	0.05	2.9957	1.8209	0.2880	0.0000	0.1214	0.1214	0.1214
26	25	0.04	3.2189	1.9569	0.3217	0.0000	0.1136	0.1136	0.1136
27	50	0.02	3.9120	2.3516	0.4250	0.0000	0.0945	0.0945	0.0945
28	100	0.01	4.6052	2.7124	0.5256	0.0000	0.0810	0.0810	0.0810

Fig. 3 Estimation of parameters (GEVM-MOM) and goodness of fit measures for gauging station Villalba, Mexico

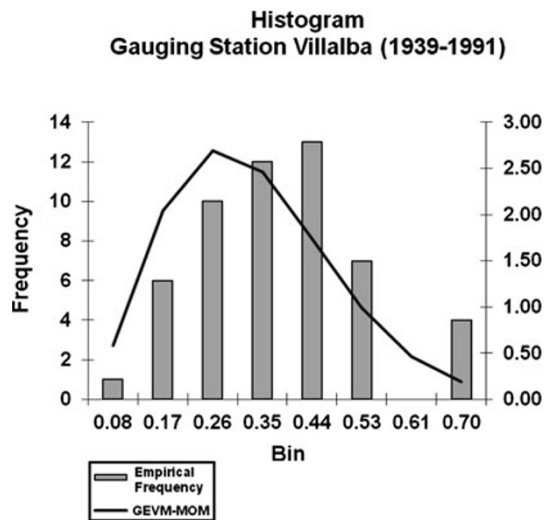


Fig. 4 Histogram and theoretical probability density function for gauging station Villalba, Mexico

provided by Excel® (Excel is a registered trademark of Microsoft Corporation, Inc.)

Discussion of results

The easy use of proposed methodology has been shown by the development of the numerical example. Using the

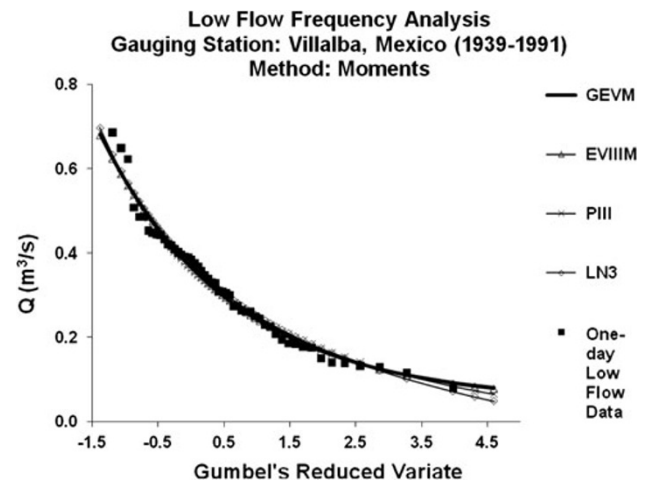


Fig. 5 Empirical and theoretical frequency curves for several models applied to 1-day low flow data at gauging station Villalba, Mexico

common spreadsheets framework provided by Excel® (Excel is a registered trademark of Microsoft Corporation, Inc.), the user has all the time on sight the formulas and results and a possible error could be spotted very easily.

The tables shown in Fig. 3 contain all the required results for a low flow frequency analysis study for a particular set of low flow data. In these tables are contained the values of the parameters, their goodness of fit measures, design values for several return periods and their confidence limits. Two

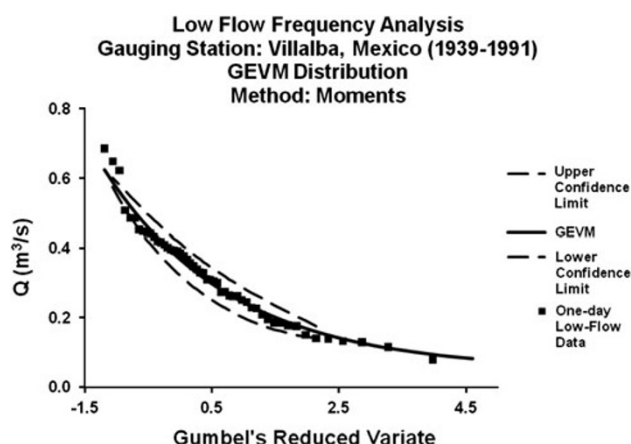


Fig. 6 Empirical and theoretical frequency curves and their confidence limits for one-day low flow data at gauging station Villalba, Mexico

different measures of goodness of fit are provided to choose among competing models.

The information contained in the graphs produced by the common spreadsheets framework provided by Excel® (Excel is a registered trademark of Microsoft Corporation, Inc.) are informative on how good is the adjustment of a particular probability distribution function to a particular set of data, this is given by the graph showing the low flow data and the adjusted model (Fig. 5), the graph that shows the theoretical probability density function and histogram of low flow data (Fig. 4) and the graph that shows the confidence limits and the adjusted model and the low flow data (Fig. 6).

Conclusions

A proposed methodology has been presented for low flow frequency analysis, using the GEVM distribution coupled with MOM method. The use of the common spreadsheets framework provided by Excel® (Excel is a registered trademark of Microsoft Corporation, Inc.) is particularly useful in education and training. The proposed methodology compares well with the existing probability distribution functions when the MOM method is applied. The straightforward application of the proposed methodology to real data, as it has been shown in example contained in the paper, makes it a versatile tool to train students or technical personnel in the field with a personal computer and a printer.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Durrans SR, Tomic S (2001) Comparison of parametric tail estimators for low flow frequency analysis. *J Am Wat Resour Assoc* 37(5):1203–1214. doi:10.1111/j.1752-1688.2001.tb03632.x
- Gumbel EJ (1958) *Statistics of extremes*. Columbia University Press, New York
- Hao Z, Singh VP (2009) Entropy-based parameter estimation for extended three-parameter Burr III distribution for low flow frequency analysis. *Trans ASABE* 52(4):1193–1202
- Iacobellis V (2008) Probabilistic model for the estimation of the T-year flow duration curves. *Wat Resour Res* 44(2), article number W02413. doi:10.1029/2006WR005400
- Jain D, Singh VP (1987) Estimating parameters of EV1 distribution for flood frequency analysis. *Wat Resour Bull* 23(1):59–71
- Kite GW (1988) *Flood and risk analyses in hydrology*. Water Resources Publications, Littleton
- Lee KS, Kim SU (2008) Identification of uncertainty in low flow frequency analysis using Bayesian MCMC method. *Hydrol Proc* 22(12):1949–1964. doi:10.1002/hyp.6778
- Matalas NC (1963) *Probability distribution of low flows*: USGS professional paper no 434-A. US Printing Office, Washington
- Offinger R (1996) *Schätzer in drei Parametrischen Weibull-Modellen und Untersuchung ihrer Eigenschaften mittels Simulation*. Diplomarbeit Institut für Mathematik der Universität Augsburg, Augsburg
- Rao R, Hamed KH (2000) *Flood frequency analysis*. CRC Press, Boca Raton
- Raynal-Villasenor JA (2010) Frequency analysis of hydrologic extremes. <http://www.lulu.com/spotlight/floodro4dot0atgmaildotcom>. Accessed 10 Sept 2011
- Raynal-Villasenor JA, Douriet-Cardenas JC (1994) Moment parameter estimators for the general extreme value distribution for the minima. *Hydrol Sci Tech J* 10(1–4):118–125
- Salas JD, Smith R (1980) *Computer programs of distribution functions in hydrology*. Colorado State University, Fort Collins
- Smakhtin V (2001) Low flow hydrology: a review. *J Hydrol* 240(3–4):147–186. doi:10.1016/S0022-1694(00)00340-1
- Taha BMJO, Charron C, St-Hilaire A (2008) Statistical model and the estimation of low flows. *Can Wat Resour J* 33(2):195–205
- Weibull W (1939a) *A statistical theory of strength of materials*: Ingeniörs Vetenskaps Akademien Handlingar, No. 151. Generalstabens Litografiska Anstalts Förlag, Stockholm
- Weibull W (1939b) *The phenomenon of rupture of solids*: Ingeniörs Vetenskaps Akademien Handlingar, No. 153. Generalstabens Litografiska Anstalts Förlag, Stockholm
- Yue S, Wang CY (2004) Scaling of Canadian low flows. *Stoch Environ Res Risk Assess* 18(5):291–305. doi:10.1007/s00477-004-0176-6